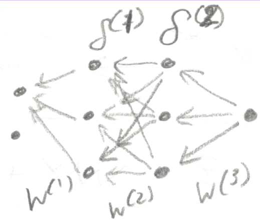
Assume $= a = \begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \end{bmatrix}$

$\uparrow$
1 sample



$$\frac{\partial J}{\partial W^{(1)}} = \begin{bmatrix} \frac{\partial J}{\partial W_{11}^{(1)}} & \frac{\partial J}{\partial W_{12}^{(1)}} \\ \frac{\partial J}{\partial W_{21}^{(1)}} & \frac{\partial J}{\partial W_{22}^{(1)}} \\ \frac{\partial J}{\partial W_{31}^{(1)}} & \frac{\partial J}{\partial W_{32}^{(1)}} \end{bmatrix}$$

$$\frac{\partial J}{\partial W_{11}^{(1)}} = \frac{\partial J}{\partial a_{11}^{(2)}} \cdot \frac{\partial a^{(2)}}{\partial z_{11}^{(2)}} \cdot \frac{\partial z_{11}^{(2)}}{\partial a_{11}^{(1)}} \cdot \frac{\partial a_{11}^{(1)}}{\partial z_{11}^{(1)}} \cdot \frac{\partial z_{11}^{(1)}}{\partial W_{11}^{(1)}}$$

$$\underbrace{= \frac{\partial z_{11}^{(1)}}{\partial W_{11}^{(1)}} = a_{11} \quad \cancel{\frac{\partial z_{21}^{(1)}}{\partial W_{11}^{(1)}} = 0} \quad \frac{\partial z_{11}}{\partial W_{12}} = a_{21}}$$

so delta coming from $z_{11}^{(1)}$ should only go into $W_{11}^{(1)}$ and $W_{12}^{(1)}$

$$\begin{bmatrix} W_{11}^{(1)} & W_{12}^{(1)} \\ \vdots & \vdots \\ \vdots & W_{32}^{(1)} \end{bmatrix} \begin{bmatrix} \delta_{11} \\ \delta_{21} \\ \delta_{31} \end{bmatrix}$$

~~we want to cascade multiply these~~

$$\begin{bmatrix} a_{11}^{(0)} \cdot \delta_{11} & a_{21}^{(0)} \delta_{11} \\ a_{11}^{(0)} \cdot \delta_{21} & a_{21}^{(0)} \delta_{21} \\ a_{11}^{(0)} \delta_{31} & a_{21}^{(0)} \delta_{31} \end{bmatrix} = (3\times1)\ (1\times2) = (3\times2) = \delta \cdot a_1^T$$

$$\delta^{(2)} = \begin{bmatrix} \delta_{11}^{(2)} \\ \delta_{21}^{(2)} \\ \delta_{31}^{(3)} \end{bmatrix}$$

$$\delta_{11}^{(1)} = \delta_{11}^{(2)} \cdot \frac{\partial z_{11}^{(2)}}{\partial a_{11}^{(1)}} + \delta_{21}^{(2)} \cdot \frac{\partial z_{21}^{(2)}}{\partial a_{11}^{(1)}} + \delta_{31}^{(2)} \frac{\partial z_{31}^{(2)}}{\partial a_{11}^{(1)}}$$

$$= \delta_{11}^{(2)} \cdot W_{11}^{(2)} + \delta_{21}^{(2)} \cdot W_{21}^{(2)} + \delta_{31}^{(2)} W_{31}^{(2)}$$

$$W^{(2)T} \underset{\text{mat mul}}{\times} \delta^{(2)}$$

*

At each step:

L = # of layers (excluding input)

$\delta^{(L)}$ = CrossEntropy Prime $\underset{\text{element}}{\times}$ $\sigma(a^{(L)})(1-\sigma(a^{(L)}))$

Update for weight: $\delta^{(L)} \cdot a_{(L-1)}^T$

Propogate error: $\delta^{(L-1)} = \left[W^{(L)T} \underset{\text{matmul}}{\times} \delta^{(L)}\right] \underset{\text{elementwise}}{\times}$ Sigmoid Prime$(z^{(L-1)})$